

04CO  
04-10-01

#2

02280

OIPE

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/808,387

DATE: 03/27/2001  
TIME: 16:04:23

Input Set : A:\seqlist.txt  
Output Set: N:\CRF3\03272001\I808387.raw

ENTERED

4 <110> APPLICANT: Kaia Palm  
5 Tonis Timmusk  
6 CeMines Research  
8 <120> TITLE OF INVENTION: MAMMALIAN NEURALIZED FAMILY OF  
9 TRANSCRIPTION REGULATORS AND USES THEREFOR  
12 <130> FILE REFERENCE: CEMRES.001A  
C--> 14 <140> CURRENT APPLICATION NUMBER: US/09/808,387  
C--> 14 <141> CURRENT FILING DATE: 2001-03-14  
14 <160> NUMBER OF SEQ ID NOS: 48  
16 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
18 <210> SEQ ID NO: 1  
19 <211> LENGTH: 1725  
20 <212> TYPE: DNA  
21 <213> ORGANISM: Homo sapien  
23 <400> SEQUENCE: 1  
24 atgggtaaca acttctccag tatccctcgt ctgccccgag gaaacccgag cgcgcgcgcg 60  
25 cggggccacc cccagaacct caaagactct atcgggggcc cttccccgt cacttctcac 120  
26 cgatgccacc acaagcagaa gcactgtccg gcagtgtctc ccagcggggg gctcccagcc 180  
27 acgccgctgc tcttccaccc gcacaccaag ggctcccaga tctcatgga cctcagccac 240  
28 aaggtgttca agaggcaggc cagcttctgc aacgccatca cttcagcaa ccgcccggtc 300  
29 ctcatctacg agcaagtcag gctgaagatc accaagaagc agtgctgctg gagcggggcc 360  
30 ctgcccgtgg gcttcaccag caaggaccgc tcccgcctcc accctgactc gctgcccagg 420  
31 tacgcctgcc ccgacctggg gtcccagagt ggcttctggg ccaaggcgtc gctgaggagg 480  
32 ttggccaatg agggcaacat catcgcatc tgggtggaca agaaggccg tgtcttcac 540  
33 cgcatacaac actcggctgt tatcgtgttc ttcagcgggg tccgcacggc cgaccgcgtc 600  
34 tggggcctgg tggacgtcta cggcctcacg cggggcgtcc agctgcttga tagcgagctg 660  
35 gtgtccccg actgtctgcg gccgcgtccc ttcaccgcc tgcggcggcc gtcgtgcgg 720  
36 cgcgaggcgg acgacgcgcg cctctcggtg agcctatgcg acctcaacgt gccgggcgcg 780  
37 gacggcgacg aggcgcgcgc ggcgcgcggc tgccccatcc cgcagaactc actcaactcg 840  
38 cagcacagcc gcgcgtgcc ggcgcagctc gacggcgacc tgcgtttcca cgcctgcgc 900  
39 gccggcgcgc acgtccgcat cctcgacgag cagacggtgg cgcgcgtgga gcacgggcgc 960  
40 gacgagcgcg cgtcgtctt caccagccgg cccgtgcgcg tggccgagac catcttcgtc 1020  
41 aaggtcacgc gctcgggtgg cgcgcggccc ggcgcgtgt cgttcggcgt caccacgtgc 1080  
42 gaccccgcca cgtgcggcc ggccgacctg ctttcagcc ctgaggccct ggtggaccgc 1140  
43 aaggaattct gggccgtgtg ccgcgtgccc gggccccctg acagcggcga catcctgggc 1200  
44 ctggtggtca acgcccagcg cagctgcac ctcagccaca atggcgcggc cgcgggcagt 1260  
45 cagctgtgcg tggacgcctc gcagccgctt tggatgtct tggcctgca cgggaccatc 1320  
46 acgcagatcc gcatactcgg ctccactatc ctggccgagc ggggtatccc gtcactcccc 1380  
47 tgcctccctg cctccacgcc aacctcgccc agtgccctgg gcagccgctt gtctgacccc 1440  
48 ttgtcagca cgtgcagctc tggccctctg ggtagctctg ctggtgggac agccccaat 1500  
49 tcgccagtga gcctgccga gtcgccagt accccaggtc tgggccaagt gagcgatgag 1560  
50 tgcaccattt gctatgaaca cgcggtggac acggtcatct acacatgtgg ccacatgtgc 1620  
51 ctctgtacg cctgtggcct gcgcctcaag aaggtctgc acgcctgctg ccccatctgc 1680  
52 cgcgcgcaca tcaaggacat catcaagacc taccgcagct cctag 1725  
54 <210> SEQ ID NO: 2  
55 <211> LENGTH: 574  
56 <212> TYPE: PRT

## RAW SEQUENCE LISTING

DATE: 03/27/2001

PATENT APPLICATION: US/09/808,387

TIME: 16:04:23

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\03272001\I808387.raw

```

57 <213> ORGANISM: Homo sapien
59 <400> SEQUENCE: 2
60 Met Gly Asn Asn Phe Ser Ser Ile Pro Ser Leu Pro Arg Gly Asn Pro
61 1 5 10 15
62 Ser Arg Ala Pro Arg Gly His Pro Gln Asn Leu Lys Asp Ser Ile Gly
63 20 25 30
64 Gly Pro Phe Pro Val Thr Ser His Arg Cys His His Lys Gln Lys His
65 35 40 45
66 Cys Pro Ala Val Leu Pro Ser Gly Gly Leu Pro Ala Thr Pro Leu Leu
67 50 55 60
68 Phe His Pro His Thr Lys Gly Ser Gln Ile Leu Met Asp Leu Ser His
69 65 70 75 80
70 Lys Ala Val Lys Arg Gln Ala Ser Phe Cys Asn Ala Ile Thr Phe Ser
71 85 90 95
72 Asn Arg Pro Val Leu Ile Tyr Glu Gln Val Arg Leu Lys Ile Thr Lys
73 100 105 110
74 Lys Gln Cys Cys Trp Ser Gly Ala Leu Arg Leu Gly Phe Thr Ser Lys
75 115 120 125
76 Asp Pro Ser Arg Ile His Pro Asp Ser Leu Pro Lys Tyr Ala Cys Pro
77 130 135 140
78 Asp Leu Val Ser Gln Ser Gly Phe Trp Ala Lys Ala Leu Pro Glu Glu
79 145 150 155 160
80 Phe Ala Asn Glu Gly Asn Ile Ile Ala Phe Trp Val Asp Lys Lys Gly
81 165 170 175
82 Arg Val Phe His Arg Ile Asn Asp Ser Ala Val Met Leu Phe Phe Ser
83 180 185 190
84 Gly Val Arg Thr Ala Asp Pro Leu Trp Ala Leu Val Asp Val Tyr Gly
85 195 200 205
86 Leu Thr Arg Gly Val Gln Leu Asp Ser Glu Leu Val Leu Pro Asp
87 210 215 220
88 Cys Leu Arg Pro Arg Ser Phe Thr Ala Leu Arg Arg Pro Ser Leu Arg
89 225 230 235 240
90 Arg Glu Ala Asp Asp Ala Arg Leu Ser Val Ser Leu Cys Asp Leu Asn
91 245 250 255
92 Val Pro Gly Ala Asp Gly Asp Glu Ala Ala Pro Ala Ala Gly Cys Pro
93 260 265 270
94 Ile Pro Gln Asn Ser Leu Asn Ser Gln His Ser Arg Ala Leu Pro Ala
95 275 280 285
96 Gln Leu Asp Gly Asp Leu Arg Phe His Ala Leu Arg Ala Gly Ala His
97 290 295 300
98 Val Arg Ile Leu Asp Glu Gln Thr Val Ala Arg Val Glu His Gly Arg
99 305 310 315 320
100 Asp Glu Arg Ala Leu Val Phe Thr Ser Arg Pro Val Arg Val Ala Glu
101 325 330 335
102 Thr Ile Phe Val Lys Val Thr Arg Ser Gly Gly Ala Arg Pro Gly Ala
103 340 345 350
104 Leu Ser Phe Gly Val Thr Thr Cys Asp Pro Gly Thr Leu Arg Pro Ala
105 355 360 365
106 Asp Leu Pro Phe Ser Pro Glu Ala Leu Val Asp Arg Lys Glu Phe Trp

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/808,387

DATE: 03/27/2001

TIME: 16:04:23

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\03272001\I808387.raw

```

107      370      375      380
108 Ala Val Cys Arg Val Pro Gly Pro Leu His Ser Gly Asp Ile Leu Gly
109 385      390      395      400
110 Leu Val Val Asn Ala Asp Gly Glu Leu His Leu Ser His Asn Gly Ala
111      405      410      415
112 Ala Ala Gly Met Gln Leu Cys Val Asp Ala Ser Gln Pro Leu Trp Met
113      420      425      430
114 Leu Phe Gly Leu His Gly Thr Ile Thr Gln Ile Arg Ile Leu Gly Ser
115      435      440      445
116 Thr Ile Leu Ala Glu Arg Gly Ile Pro Ser Leu Pro Cys Ser Pro Ala
117      450      455      460
118 Ser Thr Pro Thr Ser Pro Ser Ala Leu Gly Ser Arg Leu Ser Asp Pro
119 465      470      475      480
120 Leu Leu Ser Thr Cys Ser Ser Gly Pro Leu Gly Ser Ser Ala Gly Gly
121      485      490      495
122 Thr Ala Pro Asn Ser Pro Val Ser Leu Pro Glu Ser Pro Val Thr Pro
123      500      505      510
124 Gly Leu Gly Gln Trp Ser Asp Glu Cys Thr Ile Cys Tyr Glu His Ala
125      515      520      525
126 Val Asp Thr Val Ile Tyr Thr Cys Gly His Met Cys Leu Cys Tyr Ala
127      530      535      540
128 Cys Gly Leu Arg Leu Lys Lys Ala Leu His Ala Cys Cys Pro Ile Cys
129 545      550      555      560
130 Arg Arg Pro Ile Lys Asp Ile Ile Lys Thr Tyr Arg Ser Ser
131      565      570
134 <210> SEQ ID NO: 3
135 <211> LENGTH: 1674
136 <212> TYPE: DNA
137 <213> ORGANISM: Homo sapien
139 <400> SEQUENCE: 3
140 atgggggggac agatcacccg gagcactctc cagcactcta tcggggggccc cttccccgtc 60
141 actttctcacc gatgccacca caagcagaag cactgtccgg cagtgtgcc cagcgggggg 120
142 ctcccagcca cgccgctgct cttccacccg cacaccaagg gctcccagat cctcatggac 180
143 ctccagccaca aggctgtcaa gaggcaggcc agcttctgca acgccatcac cttcagcaac 240
144 cgccccgtcc tcatctacga gcaagtcagg ctgaagatca ccaagaagca gtgctgctgg 300
145 agcgggggccc tgcggctggg cttcaccagc aaggaccggt cccgcatcca cctgactcgc 360
146 ctgcccgaagt acgcctgccc cgacctgggt tcccagagtg gcttctgggc caaggcgctg 420
147 cctgaggagt ttgccaatga gggcaacatc atcgattctt ggggtggaaa gaagggccgt 480
148 gtcttccacc gcatcaacga ctcggtgttt atgtgtttct tcagcggggg ccgcacggcc 540
149 gacccgctct gggccctggt ggacgtctac ggcctcacgc ggggcgtcca gctgcttgat 600
150 agcgagctgg tgctcccgga ctgtctgcgg ccgcgctcct tcaccgccct gcggcgccgc 660
151 tcgctgcggc gcgaggcgga cgacgcgcgc ctctcggtga gcctatgcca cctcaacgtg 720
152 ccgggcgcgg acggcgacga ggccgcgcgc gccgcgggt gcccatccc gcagaactca 780
153 ctcaactcgc agcacagccg cgcgtgcgg gcgcagctcg acggcgacct gcgtttccac 840
154 gccctgcgcg ccggcgcgca cgtccgcata ctgcacgagc agacggtggc gcgcgtggag 900
155 caggggcgcg acgagcgcg ctcgtcttcc accagccggc ccgtgcgcgt ggccgagacc 960
156 atcttcgtca aggtcacgcg ctcggttggc gcgcggccc gcgcgctgtc gttcggcgctc 1020
157 accacgtgcg accccggcac gctgcggccg gccgacctgc ctttcagccc tgaggccctg 1080
158 gtggaccgca aggaattctg ggccgtgtgc cgcgtgccc ggcccctgca cagcggcgac 1140

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/808,387

DATE: 03/27/2001

TIME: 16:04:23

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\03272001\I808387.raw

```

159 atcctgggcc tgggtgtcaa cgccgacggc gagctgcacc tcagccacaa tggcgcggcc 1200
160 gccggcatgc agctgtgctg ggacgectcg cagccgcttt ggatgctctt cggcctgcac 1260
161 gggaccatca cgcagatccg cctcctcggc tccactatcc tggccgagcg gggatatccg 1320
162 tcactccccct gctccccctgc ctccacgcca acctcgccca gtgccctggg cagccgcctg 1380
163 tctgaccctt tgtcagcac gtgcagctct ggccctctgg gtagctctgc tgggtgggaca 1440
164 gcccccaatt cgccagttag cctgcccag tgcagtgga cccaggtct gggccagtgg 1500
165 agcgatgagt gcaccatttg ctatgaacac gcggtggaca cggatcatcta cacatgtggc 1560
166 cacatgtgcc tctgctacgc ctgtggcctg cgcctcaaga aggtcttgca cgcctgctgc 1620
167 cccatctgcc gccgccccat caaggacatc atcaagacct accgcagctc ctag 1674
169 <210> SEQ ID NO: 4
170 <211> LENGTH: 557
171 <212> TYPE: PRT
172 <213> ORGANISM: Homo sapien
174 <400> SEQUENCE: 4
175 Met Gly Gly Gln Ile Thr Arg Ser Thr Leu His Asp Ser Ile Gly Gly
176 1 5 10 15
177 Pro Phe Pro Val Thr Ser His Arg Cys His His Lys Gln Lys His Cys
178 20 25 30
179 Pro Ala Val Leu Pro Ser Gly Gly Leu Pro Ala Thr Pro Leu Leu Phe
180 35 40 45
181 His Pro His Thr Lys Gly Ser Gln Ile Leu Met Asp Leu Ser His Lys
182 50 55 60
183 Ala Val Lys Arg Gln Ala Ser Phe Cys Asn Ala Ile Thr Phe Ser Asn
184 65 70 75 80
185 Arg Pro Val Leu Ile Tyr Glu Gln Val Arg Leu Lys Ile Thr Lys Lys
186 85 90 95
187 Gln Cys Cys Trp Ser Gly Ala Leu Arg Leu Gly Phe Thr Ser Lys Asp
188 100 105 110
189 Pro Ser Arg Ile His Pro Asp Ser Leu Pro Lys Tyr Ala Cys Pro Asp
190 115 120 125
191 Leu Val Ser Gln Ser Gly Phe Trp Ala Lys Ala Leu Pro Glu Glu Phe
192 130 135 140
193 Ala Asn Glu Gly Asn Ile Ile Ala Phe Trp Val Asp Lys Lys Gly Arg
194 145 150 155 160
195 Val Phe His Arg Ile Asn Asp Ser Ala Val Met Leu Phe Phe Ser Gly
196 165 170 175
197 Val Arg Thr Ala Asp Pro Leu Trp Ala Leu Val Asp Val Tyr Gly Leu
198 180 185 190
199 Thr Arg Gly Val Gln Leu Leu Asp Ser Glu Leu Val Leu Pro Asp Cys
200 195 200 205
201 Leu Arg Pro Arg Ser Phe Thr Ala Leu Arg Arg Pro Ser Leu Arg Arg
202 210 215 220
203 Glu Ala Asp Asp Ala Arg Leu Ser Val Ser Leu Cys Asp Leu Asn Val
204 225 230 235 240
205 Pro Gly Ala Asp Gly Asp Glu Ala Ala Pro Ala Ala Gly Cys Pro Ile
206 245 250 255
207 Pro Gln Asn Ser Leu Asn Ser Gln His Ser Arg Ala Leu Pro Ala Gln
208 260 265 270
209 Leu Asp Gly Asp Leu Arg Phe His Ala Leu Arg Ala Gly Ala His Val

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/808,387

DATE: 03/27/2001

TIME: 16:04:23

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\03272001\I808387.raw

```

210          275          280          285
211 Arg Ile Leu Asp Glu Gln Thr Val Ala Arg Val Glu His Gly Arg Asp
212          290          295          300
213 Glu Arg Ala Leu Val Phe Thr Ser Arg Pro Val Arg Val Ala Glu Thr
214 305          310          315          320
215 Ile Phe Val Lys Val Thr Arg Ser Gly Gly Ala Arg Pro Gly Ala Leu
216          325          330          335
217 Ser Phe Gly Val Thr Thr Cys Asp Pro Gly Thr Leu Arg Pro Ala Asp
218          340          345          350
219 Leu Pro Phe Ser Pro Glu Ala Leu Val Asp Arg Lys Glu Phe Trp Ala
220          355          360          365
221 Val Cys Arg Val Pro Gly Pro Leu His Ser Gly Asp Ile Leu Gly Leu
222          370          375          380
223 Val Val Asn Ala Asp Gly Glu Leu His Leu Ser His Asn Gly Ala Ala
224 385          390          395          400
225 Ala Gly Met Gln Leu Cys Val Asp Ala Ser Gln Pro Leu Trp Met Leu
226          405          410          415
227 Phe Gly Leu His Gly Thr Ile Thr Gln Ile Arg Ile Leu Gly Ser Thr
228          420          425          430
229 Ile Leu Ala Glu Arg Gly Ile Pro Ser Leu Pro Cys Ser Pro Ala Ser
230          435          440          445
231 Thr Pro Thr Ser Pro Ser Ala Leu Gly Ser Arg Leu Ser Asp Pro Leu
232          450          455          460
233 Leu Ser Thr Cys Ser Ser Gly Pro Leu Gly Ser Ser Ala Gly Gly Thr
234 465          470          475          480
235 Ala Pro Asn Ser Pro Val Ser Leu Pro Glu Ser Pro Val Thr Pro Gly
236          485          490          495
237 Leu Gly Gln Trp Ser Asp Glu Cys Thr Ile Cys Tyr Glu His Ala Val
238          500          505          510
239 Asp Thr Val Ile Tyr Thr Cys Gly His Met Cys Leu Cys Tyr Ala Cys
240          515          520          525
241 Gly Leu Arg Leu Lys Lys Ala Leu His Ala Cys Cys Pro Ile Cys Arg
242          530          535          540
243 Arg Pro Ile Lys Asp Ile Ile Lys Thr Tyr Arg Ser Ser
244 545          550          555

```

247 &lt;210&gt; SEQ ID NO: 5

248 &lt;211&gt; LENGTH: 1161

249 &lt;212&gt; TYPE: DNA

250 &lt;213&gt; ORGANISM: Homo sapien

252 &lt;400&gt; SEQUENCE: 5

```

253 atgggtaaca acttctccag tatccctcgt ctgccccgag gaaaccgag ccgcgcgcgc 60
254 cggggccacc ccagaaacct caaagatagc gagctggtgc tcccggactg tctgcggccg 120
255 cgctccttca ccgccctgcg gggccgtcgt ctgcggcgcg aggcggacga cgcgcgcctc 180
256 tcggtgagcc tatgcgacct caacgtgccg ggcgcggacg gcgacgaggg ccgcgccggc 240
257 gccggctgcc ccatcccgca gaactcactc aactcgcagc acagccgcgc gctgccggcg 300
258 cagctcgacg gcgacctgcg ttccacgccc ctgcgcgcgc gcgcgcacgt ccgcacctc 360
259 gacgagcaga cggtagcgcg cgtggagcac ggcgcgcagc agcgcgcgct cgtcttcacc 420
260 agcgggcccg tgccgctggc cgagaccatc ttctcaagg tcacgcgctc gggtagcgcg 480
261 cggcccgcg cgctgtcgtt cggcgtcacc acgtgcgacc ccggcacgct gcggccggcc 540

```

PJI:

**Please Note:**

Use f n and/ r Xaa have been detected in the Sequence Listing. Please review the Sequence Listing t ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

VERIFICATION SUMMARY                      DATE: 03/27/2001  
PATENT APPLICATION:    US/09/808,387        TIME: 16:04:24

Input Set : A:\seqlist.txt  
Output Set: N:\CRF3\03272001\I808387.raw

L:14 M:270 C: Current Application Number differs, Replaced Current Application No  
L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:1922 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:46  
L:1939 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:47  
L:1941 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:47